Volume 0 (1981), Number 0 pp. 1–12 COMPUTER GRAPHICS *forum*

Olfaction and Selective-Rendering

Carlo Harvey[1], Thomas Bashford-Rogers[2], Kurt Debattista[3], Efstratios Doukakis[3] and Alan Chalmers[3]

[1]Birmingham City University, School of Computing and Digital Technology, UK.
[2]UWE Bristol, Computer Science and Creative Technologies, UK.
[3]University of Warwick, Warwick Manufacturing Group, UK.

Abstract

Accurate simulation of all the senses in virtual environments is a computationally expensive task. Visual saliency models have been used to improve computational performance for rendered content, but this is insufficient for multi-modal environments. This paper considers cross-modal perception and, in particular, if and how olfaction affects visual attention. Two experiments are presented in this paper. Firstly, eye tracking is gathered from a number of participants to gain an impression about where and how they view virtual objects when smell is introduced compared to an odourless condition. Based on the results of this experiment a new type of saliency map in a selective-rendering pipeline is presented. A second experiment validates this approach, and demonstrates that participants rank images as better quality, when compared to a reference, for the same rendering budget.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Viewing Algorithms I.4.8 [Computer Graphics]: Image Processing and Computer Vision—Scene Analysis - Object Recognition I.4.8 [Computer Graphics]: Image Processing and Computer Vision—Scene Analysis - Tracking

Keywords: Multi-Modal, Cross-Modal, Saliency, Olfaction, Graphics, Selective-Rendering

1. Introduction

A major research challenge of Virtual Environments (VEs) is to accurately simulate real world environments. This is motivated by the increasing use of VEs in a wide range of applications such as concert hall and architectural design [Dal, Nay93], simulation and training, and immersive video games [MBT*07, RLC*07, GBW*09]. Multisensory VEs aim to deliver more sensory information and yield an increased sense of immersion and accuracy over single modality environments [DM95], but multisensory VEs have inherent perceptual affects that have to be understood [AMMG16]. Furthermore, such multisensory VEs can aid object recognition and placement; identification and localisation; and generating conclusions pertaining to the scale and shape of the environment [Bla97].

Multisensory VEs can be very computationally complex due to having to simulate all senses [GTLG14], however, limitations of the human sensory system can be used in order to improve the performance of rendering systems by computing certain senses at lesser quality. Examples of this being used to decrease the auditory [TGD04, MBT*07] or visual [RCHR07, CCL02, RFWB07, RBF08] rendering complexity with little or no perceivable quality difference to a user have been implemented and verified. Moreover, it has been shown that it is possible to increase the perceptual quality of a stimulus in one modality by directing gaze due to the introduction of another modality [MDCT05]. This can be used for improving the perception of a material's quality [BSVDD10] or for Level-of-Detail (LOD) selection [BCB*09b, GBW*09, RKH*13].

1.1. Smell and Attention

The sense of smell, olfaction, is a major sense of humans. Smell has been linked with influencing mood, emotion, memory, social behaviour and even partner choice [Jac07]. Despite this, smell has largely been omitted from VEs, waiving the role it plays in human behaviour.

It has been well established that visual cues have a marked effect on improving olfactory performance. Zellner [ZBE91] demonstrated that odours matched appropriately with a colour were rated more pleasant than inappropriately matched odours. Sakai *et al.* [SIS*05] showed that watching congruent pictures had the effect of increasing the pleasantness and also odour intensity as opposed to incongruent picture matching. Seignuric *et al.* [SDJ*10] explored the influence of prior learned associations between an odour and a visual stimulus naturally associated with that odour on eye saccades and fixations. They showed that the odour-related visual cue was explored faster and for a shorter time in the presence of a congruent odour. Harvey *et al.* [HBRDC11] introduced an eye-tracking experiment that was normalised by visual saliency to show that in the presence of congruent olfactory cues conventional saliency maps for vision in the spatial domain no longer can be relied upon in the same way. Chen *et al.* [CZC*13] later performed a similar study to further corroborate this effect stating, "Our discoveries provide robust empirical evidence for a multimodal saliency map that weighs not only visual but also olfactory inputs."

Some research has investigated the influence of olfactory cues on visual attention. For a recent overview of bottom-up and top-down attention, see the work by Katsuki and Constantinidis [KC14]. One study presented by Millot *et al.* has shown smell can impact participant behaviour towards visuals [MBM02], in that ambient smell produced a faster response to a visual stimuli in a sensory-motor task than a condition with no ambient smell. Knasko [Kna95] showed participants looked longer at slides in the presence of a pleasant odour such as baby powder and chocolate than with no odour present. Seo *et al.* [SRMN10] extended this by showing a distinct effect of congruency upon viewing time and in addition explored where eyes fixated, given congruent and incongruent images as a visual stimulus.

While prior research investigates perceptual effects, in every instance this ignores the effect of visual saliency on the study and does not consider the application to the field of computer graphics. Whilst the work implies a conditioning to smell congruencies it ignores the possibility results could be inspired by a biological predisposition under *bottom-up* conditions to spend more time attending that particular congruent feature presented than another slide presented asynchronously. We investigate this possibility and employ the findings into a rendering pipeline.

In this paper we propose a general algorithm to combine olfactory saliency with visual saliency in the spatio-visual domain. In particular, this paper considers the effect of introducing an olfactory stimulus to a VE, on a user's visual attention towards imagery. Based on eye tracking data from an experiment, new saliency maps are derived. These are used to reduce render times while maintaining perceptual equality. This is validated in a further experiment. Specifically, we make the following contributions:

- Quantification of the area of influence and weighting strategies to guide attention in images when a congruent smell is present.
- A general framework to generate olfactory saliency maps.
- Combination of the traditional visual saliency map with the olfactory saliency map into a multi-modal saliency map. This is used to reduce rendering time in this paper. The technique could be more broadly applied to any olfactory-visual interface.
- An experiment which validates the use of the multi-modal maps to reduce rendering time, but maintaining similar perceptual quality to reference images computed at higher sampling rates.

2. Experiment 1: Data Capture

The effect of olfaction on saliency combined with *bottom-up* visual attention has not been investigated. In order to create a rendering framework; both need to be considered for VEs. The aim of this initial experiment is to investigate the effect of smell congruencies on visual saliency in VEs; to identify and quantify the *bottom-up* conditions that cause more time being spent attending that particular congruent feature. This section describes the methodology of the first experiment, designed to understand whether smell affects visual attention and if visually salient objects affect this process also. The data capture and results of the experiment are used to develop a novel selective-rendering approach which addresses the need for a framework and a more general heuristic for the work to be more readily applicable to image rendering.

2.1. Design

The experiment design required participants to view a series of visual stimuli while being exposed to an odour condition. Since there are results [SRMN10,SDJ*10,ZBE91,Kna95,HBRDC11] that suggest a link between visual attention and congruent smells, a replication of the result shown in [SDJ*10] was considered and augmented with a study of what role visual saliency plays in this observed effect. The work presented in this paper extends and further builds upon this augmentation by Harvey *et al.* [HBRDC11] by re-analysing the findings (Section 3.2) and subsequently evaluating the suggested approach (Section 5).

The experiment presents four concurrent stimuli in a slide, one in each quadrant. Three of the four stimuli (apple, banana, orange), have a congruent smell associated with them while the fourth (liquorice or strawberry termed *variable*) does not.

Two independent variables were used in a 2 (*saliency*) $\times$ 3 (*smell*) design. The first, *saliency*, denotes whether one of the visual stimuli was more salient than the others in the presented stimuli. To achieve this the two sets of stimuli were created, one where *variable* was set to liquorice, the non-salient condition, and, the salient condition consisting of a strawberry replacing the liquorice. The images were specifically chosen to have similar saliency besides the *variable* factor between the two sets. *saliency* was a between-participants variable in order to avoid participants having to repeat the experiment with only slight modifications.

The second independent variable was the odour condition *smell* which was composed of three possible states no smell, congruent smell and non-congruent smell. *smell* was a within-participants variable in which all participants experienced three smells, one for each of the related stimuli, apple, banana, and orange and a non-smell case.

The dependent variable was the time the eye tracker detected that the participant spent in the quadrant corresponding to one of the visual stimuli.

A number of hypotheses were considered based on the literature. $\mathbf{H}_a$**:** the saliency and non-saliency conditions would produce different results and the object which is more visually salient would have an effect on visual attention. $\mathbf{H}_b$**:** the stimuli associated with the congruent smell would be the one most attended to.

2.2. Materials

2.2.1. Visual Saliency Choice

Predictors of visual attention are used throughout this work for setting the *saliency* variable in this experiment and, subsequently, used in the design and evaluation of the selective renderer. The Itti and Koch [IKN98] visual attention saliency model was chosen as it has been used as an accepted benchmark of visual prediction metrics in the field of VEs for a number of years. The implementation of this algorithm used throughout was that of the freely available open source code developed by the iLab Neuromorphic Vision Laboratory of University Southern California [iLa].

2.2.2. Stimuli

The experiment used two sets of images, for each of the two conditions in the *saliency* variables. One set had an image with higher visual saliency and the other set had an image with a lower saliency; strawberry and liquorice respectively.

The other images for the remaining three quadrants were constant between sets: apple, banana and orange. The four images from each image set was displayed on the screen in random quadrants. Images for a slide were assigned a random quadrant until all images for the slide were assigned a unique quadrant. For a slide set (four slides), this process was repeated four times. One possible combination for a higher saliency image set slide is shown in Figure 2. The reason for the randomness was in an attempt to mitigate the experiment learning effect whereby participants start to be able to predict the slides to be delivered.

Besides the *variable* factor between the two sets, the images had similar saliency. As one can see from the saliency map of the slide in Figure 2; the strawberry comes out on top in a winner-takes-all combination of the local intensity, colour and orientation feature maps. A basic quantitative measure, S_V, was used to represent the saliency of the images. This entailed a heuristic of summing the normalised saliency map pixel intensity values for the relevant slide quadrant and dividing this by the area of pixel space occupied in the quadrant by values greater than zero (i.e. non-white space), this is shown by Equation 1.

$$S_V = \frac{1}{i}\sum_{x,y=0}^{x=w,y=h} I_V(x,y) \quad \text{where,} \quad i = \sum_{x,y=0}^{x=w,y=h} \mathbb{1}_{I_V(x,y)>0} \tag{1}$$

where I_V is the image matrix for the visual saliency operator, x,y are pixel coordinates and w,h are the width and height of the image respectively.

2.2.3. Equipment

An eye tracker was used to record the participants' eye movements. The model used was a faceLAB 5 by Seeing Machines; a passive measuring device with no extraneous materials connected to the participant. This system provided real time blink, saccade and fixation estimates.

The smell delivery system used was a Vortex Activ produced by Dale Air [Act10]. This device has four fan emission chambers which are programmable via a micro controller to guide chamber impulse onset, duration and which chamber in the device to fire. Polytetrafluoroethylene (PTFE) tubes were used to minimise the effect of adsorption of smell molecules onto the surface of the inner tubing which could bias subsequent data collection as participants went through the experiment. Tubes were blown clear before and after each participant for a period of 10 minutes, as well as an empty chamber pumping air during the experiment when a smell chamber was not firing. In addition to this, during experiment down time, room fans and a room neutral deodorant were used in order to minimise discrepancies between subsequent datasets. The air conditioning in the room acted to ventilate the area so as not to start a subsequent experiment in a contaminated environment. A Neutradol room neutral deodoriser was left in the room overnight so any lingering molecules would bind with the substrate and not linger on surfaces that could be excited during new experiment setup. In addition, this negated the minimal effect the passive air flow in the room could have on exciting the emission chambers. The monitor used for display was a 37" LCD display.

2.2.4. Setup

Experimental setup and placement is shown in Figure 1. The participant was sat on a chair, with the backrest of the chair 115cm from the display. One metre long PTFE tubes were drawn from the device chambers and clipped to the top of a participant's collar. The monitor was set at a 1280×1024 resolution (images displayed corresponded to this resolution). The distance between equipment was constant between participants and controlled as such.

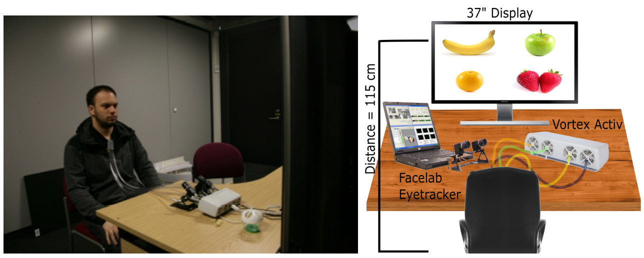

Figure 1: (l): Experimental setup photograph. (r): Figurative experimental setup schema.

2.3. Procedure

Participants were told the following, "You are asked to free-view the image presented on the screen in front of you. Your eye movements will be recorded during the period in which you are viewing the screen."

Each participant was randomly assigned to one image set and was presented four one minute slides and four 30-second buffer slides for nose desensitisation, for a total of a six minute experiment. The buffer slides displayed a small red circle in the centre of the screen to minimise discrepancies in gaze quadrant analysis and to create a fixation in a free area so free-view conditions are started without bias when the next main slide starts. The four main slides cycled randomly through the subset of variables; i.e. which smell to emit alongside the display of the slide. Smells were delivered to the participant for the full duration of the slide. Eye tracking occurred throughout participants free viewing the slides under these conditions. Upon termination participants were debriefed on the goals of the experiment.

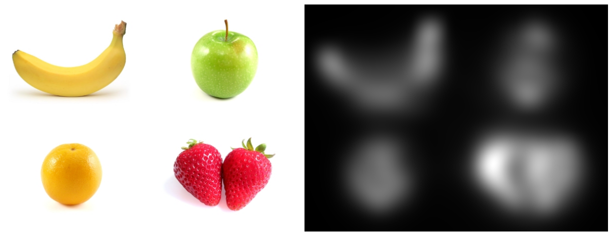

Figure 2: (l): A possible combination of images for a slide presented to a participant. Slide from the salient image set. (r): Saliency map for the slide shown.

2.4. Participants

A total of 30 participants took part in this experiment, 21 males and 9 females. 15 were assigned to the salient image set and 15 to the non-salient image set. Participants had an average age of 26.7. A participant selection criteria for progression to the study was undertaken, the participant had to indicate that they had normal or corrected to normal vision at the time of testing. In addition, the participants had to report no anosmia. Participation in this study was voluntary, recruitment to this study was facilitated through email and the sample was drawn from staff at students at a UK university.

3. Results

This section provides an overview of the timings and provides statistical analysis of the results.

3.1. Timings

The average time spent in a quad region for each smell condition is shown in Table 1 and Table 2 for saliency and no saliency conditions respectively. Not every column total will sum to 60s, the duration of a slide, due to lost tracking data or time spent gazing outside slide screen boundaries. However, time spent within the salient region decreases when the smell condition is presented asynchronously with a congruent object. In addition time spent viewing the congruent object with the smell conditions is increased drastically under both image sets. When no-smell condition is present there is no smell congruency; as such standard visual saliency predictive metrics prevail and the strawberry is most viewed. As can be seen in the control column of Table 2, the average time is balanced between quadrants.

	Smell			No Smell
	Apple	Banana	Orange	Control
Apple	**21.13**	6.04	6.34	9.37
Banana	6.26	**26.34**	5.71	6.69
Orange	6.88	6.96	**23.6**	6.89
Strawberry	13.31	11.91	11.99	**23.91**

Table 1: Average gaze time (in s) within each quadrant for the salient image set under each odour condition. Largest values per smell condition emboldened.

	Smell			No Smell
	Apple	Banana	Orange	Control
Apple	**20.38**	9.29	11.9	10.22
Banana	11.45	**23.2**	9.11	**14.18**
Orange	11.24	9.24	**20.81**	12.10
Liquorice	10.9	9.90	9.92	13.60

Table 2: Average gaze time (in s) within each quadrant for the non-salient image set under each odour condition. Largest values per smell condition emboldened.

3.2. Statistical Analysis

Analysis is conducted via a repeated measure ANOVA via a 2 (*saliency*) $\times$ 3 (*smell*) factorial design. In order to analyse the data the congruent and non-congruent smell conditions are collapsed to enable direct comparison with the no-smell condition.

3.2.1. Overall Results

The main effect of *saliency* was significant $F(1, 28) = 5.38$, $p < 0.05$, indicating a difference in results across the two groups. The salient condition was attended to for an average of 18.62s compared to the non-salient at 15.48s across all other conditions, for the variable quadrant. This indicates acceptance of $\mathbf{H}_a$**:** that the saliency and non-saliency conditions would produce different results with the salient object having an effect on visual attention.

The main effect of *smell* did not violate the assumption of sphericity (via Maulchy's test, $p > 0.05$) and was found to be significant $F(2, 56) = 35.6$, $p < 0.01$. The means of the three conditions were: no smell = 18.99s, congruent = 22.94s, non-congruent = 9.23s. This allows us to accept $\mathbf{H}_b$**:** that the stimuli associated with the congruent smell would be the one most attended to.

Pairwise comparisons, with Bonferroni correction, found significant differences between the non-congruent case and both the no smell and congruent case. No differences were noted between the no smell and congruent conditions. For pairwise comparisons when there is a higher saliency object as the between participant variable there is no significant difference between Congruent Smell and No Smell control conditions; indicating that a highly salient object without an identifiable smell and a normal object with a related smell are equally attractive.

The interaction of *saliency* $\times$ *smell* was also analysed. It did not violate the assumption of sphericity (via Maulchy's test, $p > 0.05$) and was considered significant $F(2,56) = 6.41$, $p < 0.05$. Figure 3 shows this interaction; it is clear that the no-smell condition is strongly affected by the saliency and no saliency condition, which produces the significant interaction. This also concurs with $\mathbf{H}_a$.

The results for the salient and non-salient conditions are analysed separately in the following.

3.2.2. Results for non-salient condition

For the non-salient condition of the *saliency* variable, the results are calculated using repeated-measures ANOVA for the smell variable.

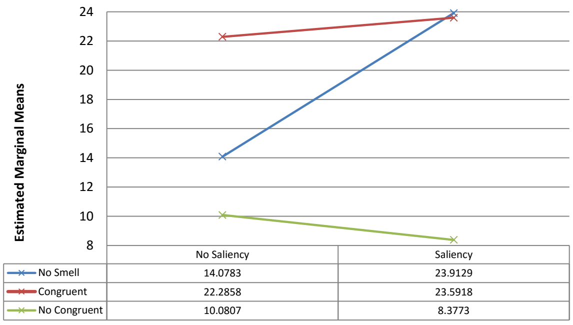

Figure 3: Interaction of *saliency* × *smell* showing the effect the salient image factor has on the no-smell condition across sets.

The assumption of sphericity was not violated (via Maulchy's test, $p > 0.05$) and the main effect of smell was significant $F(2, 28) = 17.38$, $p < 0.01$. The means of the three conditions were: no smell = 14.08s, congruent = 22.29s, non-congruent = 10.08s.

Pairwise comparisons, using Bonferroni corrections, give significant differences for congruent smell against both the other two conditions, while, there was no significant difference between the no smell and non-congruent smell conditions. This indicates that if none of the objects are strongly salient then the smell congruency dominates the visual field.

This correlates with the results reported by Seo *et. al.* [SRMN10]. They found that when participants were exposed to an olfactory stimulus, they focused significantly more often and longer at the congruent object among the four different objects being displayed, as compared to a non-odour condition.

3.2.3. Results for Salient condition

For the salient condition of the *saliency* variable, results follow a similar pattern for the non-salient condition but with a change in the pairwise comparisons discussed below. Analysis was performed via repeated-measures ANOVA for the smell variable. The assumption of sphericity was again not violated (via Maulchy's test, $p > 0.05$) and the main effect of smell was significant $F(2, 28) = 23.46$, $p < 0.01$. The means of the three conditions were: no smell = 23.91s, congruent = 23.59s, non-congruent = 8.38s.

Pairwise comparisons, with Bonferroni correction, demonstrated a significant difference between the non-congruent smell condition and the other two conditions; this is in contrast to the saliency case and further supports the results of the interaction seen in the overall results indicating that saliency and congruent smell do affect visual attention. The pairwise comparisons can be seen more clearly in Table 3.

3.3. Gaze Point Analysis

3.3.1. Density Estimation

Gaze points within the area of interest for the pertaining smell condition, after density estimation on the set of gaze points, appear to be modulated by a standard visual saliency model; as seen in Figure 4. However, on top of being modulated by visual saliency it

Condition	p, df=2			
non-saliency	<0.05	congruent	no smell	non-congruent
saliency	<0.05	no smell	congruent	non-congruent
overall	<0.05	congruent	no smell	non-congruent

Table 3: Pairwise comparisons for the various conditions. The statistical groupings are discerned by the encircling - circled groups are not significantly different.

also overrides the standard visual saliency predictor that stipulated the strawberry would have been the region of most interest; which was the case under the control condition. In combination with the results presented in Table 3 this shows a new model is needed for saliency when there are olfactory stimuli present in VEs because both saliency and congruent smells do affect visual attention. Current models do not support this perceptual interaction.

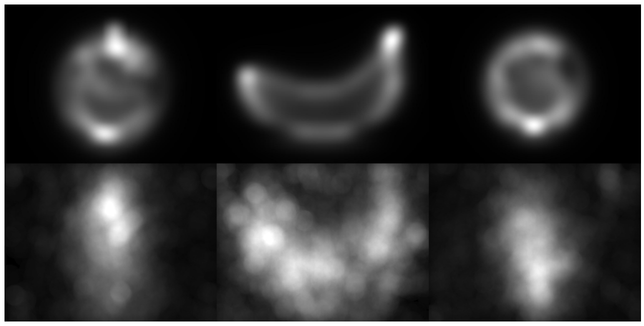

Figure 4: Top Row: Saliency maps for smell emitting images used in the slides: (l) Apple, (m) Banana, (r) Orange. Bottom Row: The corresponding density estimation of the gaze points collected on the smell emitting visual congruency. Kernel density estimation was used to perform this summation, a kernel window of $h = 50$ pixels was used in this case.

4. Selective-Rendering for Visual Attention

One of the main themes in the field of Computer Graphics is the generation of images of VEs. The classic approach to this problem is to model the VE using a selection of primitives with surface properties assigned to each primitive of the scene. In addition, it is necessary to specify the position and properties of any light sources which contribute to the radiometric quantities associated with the VE. Images of this virtual scene can then be generated by ray tracing onto the primitives of the VE through an image plane that takes into account a camera model and determining the colour of each image pixel in this plane using the assigned surface properties and the specified lighting for a given scene. This process is known as rendering.

A remodulated spatial saliency map accounting for smell congruencies within a virtual scene can be used to guide a sampling heuristic in the image plane for selectively rendering that VE.

Selective-rendering strategies are varied and diverse and are designed to address a particular aspect of the rendering pipeline. There is a need to compute certain, more relevant areas of an image, at a higher level of fidelity. These areas may be due to perception or because geometry introduces extra rendering complexity. To achieve a converging approximation of the rendering equation [Kaj86] at a certain point requires many samples and selective-rendering algorithms have been deployed successfully to this task. Many selective-rendering algorithms have been published in recent years (for a detailed overview of selective-rendering strategies see [Deb06]). One of the methods of analysing their performance for suitability, particular in the case of application to a perceptual artefact, is through psychophysical experiments to compare the algorithms against the real scenes they are purporting to depict. This section presents the method used to attenuate a conventional image saliency map based upon the observed data shown in Section 2.

4.1. Weighting Function

The recorded gaze points fit a normal distribution. Extracting a matte of a smell emitting object from the scene gives a binary predictor of the visual focus for the smell emitter in the scene. This should be modulated by the variance in attention to the smell emitter in the scene. This is done to match the normal distribution in gaze pattern observed from the measured data. The standard deviation from the measured eye track data points is used as the σ in the input to a gaussian kernel: $\frac{1}{2\pi\sigma^2} \cdot e^{-\frac{(x-\mu^2)}{2\sigma^2}}$. This gives a predictor of the area of influence of the smell attention deviation. This area of influence corresponds, from the measured eye-tracking data, to 63 pixels. From the viewing distance of 115cm and the correspondent screen space of 63 pixels the visual degrees for the σ value of this area of influence corresponds to **4.7°**.

Convolving the binary predictor of the visual focus for the smell emitter in the scene with the data-driven Gaussian kernel yields an estimator for the visual focus for the smell emitting object. This is represented as a smell matte gaussian blur, I_S, in Equation 2. Multiplying this with the original visual saliency map for the scene, I_V, and subsequently normalising gives a reattenuation factor for the original saliency for the scene. Added to the original map guided via a weighting function (Equation 2), yields a new model for visual saliency when a smell emitting object is present.

These spatial saliency maps which account for smell congruencies, termed olfactory-visual saliency maps, can be used in selective-rendering pipelines to act as a predictor for the visual attention to a scene when a smell emitter is present. The model for olfactory-visual saliency maps, I_O, is one to attenuate conventional saliency maps of a given scene, I_V, via the modulated matte, I_S.

$$I_O = (1-w)norm((max((I_V - I_S),0)) + (w)norm((I_V \circ I_S)) \quad (2)$$

where $max(\cdot)$ and $min(\cdot)$ are functions to determine the respective maximum and minimum value of the image arrays, $norm(\cdot)$ is a normalisation procedure to normalise the resultant image arrays between 0 and 1. For clarity, $\circ$ is the Hadamard product. Section 3.2 shows that when participants were exposed to an olfactory stimulus, they focused significantly more often and longer at the congruent object among the four different objects being displayed, as compared to a non-odour condition. Determining an appropriate measure for the weight w is critical to test mapping the observed findings to a perceptually based system. The weight (w) in Equation 2 is set as the ratio of the time spent looking at the smell congruency, versus the total time spent looking at the rest of the image, weighted by their respective screen real estate and saliency measure, S_V, (Equation 1). This, when averaged over all participants, gave a value $w = 0.621$.

4.2. Modal Map Generation

Figure 5 shows the rendered image of the VE, olfactory-visual saliency map, visual saliency map and the smell gaussian blurs for the Lounge, Kitchen and Restaurant scenes. The reweighting strategy adopts the pipeline outlined in Section 4.1 and Equation 2 and can be seen visually in Figure 5.

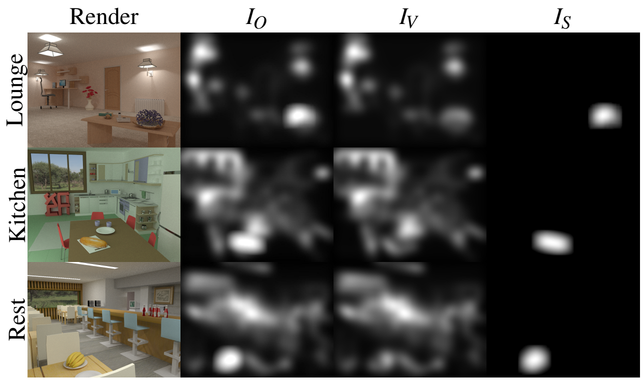

Figure 5: Render (left): Output rendered image of the virtual scene using I_O as the guide for the selective sampling strategy, olfactory-visual saliency map: I_O (centre-left), visual saliency Map: I_V (centre-right) and smell gaussian blurs: I_S (right) for the three scenes. (top to bottom): Lounge, Kitchen and Restaurant scenes.

5. Experiment 2: Application

This section presents the description of a second psychophysical experiment that aims to compare two frequently used selective-rendering operators against the algorithm reported in Section 4.1. The purpose of this investigation is to determine which technique appears closest to a reference in terms of the participant comparison judgement.

5.1. Design

The experiment was conducted on 28 participants and as such the scale of the test was not one that suited rating data [SC88]. Additionally, if participants have not been trained prior to the experiment on a series of test images the data becomes unreliable as it is not certain the participant was making informed choices or simply guessing [Ken75]. Due to the nature of the task proposed, to avoid methods such as rating or ranking which have problems associated with them [Ken75,SC88], the approach adopted was to present each subject with a pair of rendered images, A and B, in addition to a reference image, G. This allows the participant to make a decision as to which image produced with a given selective-rendering strategy is closer to G. Specifically, the participants are not presented with the A or B image concurrently, but subsequently, this is referred to as two-interval forced choice (2IFC) assessment.

Each participant was assigned 9 image sets, 3 for each scene (shown in Table 4), in random order and the A or B image was randomised within that image set. Participants were presented slides in the order A|B→G→A|B with a decision slide that waited for an input as to a decision: *"which slide A or B is closest, in your opinion, to slide G?"*. The experiment asks the participants to select which image was overall the most like a reference image (overall similarity). This technique is known as paired-comparisons. The advantage of this approach is not only simplicity, but it allows an evaluation of the transitivity, that is, the within-subject consistency of the data, as well as the between-subject consistency. The selection of a range of test scenes is of particular import to this investigation, to eliminate scene dependance.

5.2. Materials

5.2.1. Stimuli

5.2.1.1. Rendered Images The experiment used 9 image sets, 3 for each scene. The conditions tested are highlighted in Table 4.

A|B	G	A|B
Uniform	Reference	Visual Saliency
Uniform	Reference	Olfactory-Visual Saliency
Visual Saliency	Reference	Olfactory-Visual Saliency

Table 4: Smell: Experimental Conditions Tested

In total three selective-rendering algortihms were used: *uniform* (Uni), a naive algorithm that samples each pixel in the scene equally based on some overriding *rendering cost function*; *visual saliency* (Sal), as first inspired by Yee et al. [YPG01] this method uses a map of visual feature predictions to guide the pixel sampling strategy where the most samples are directed towards the area of the map deemed to be most visually important; and *olfactory-visual saliency* (SmSal), the modulation of visual saliency for when their is a congruent smell present is used to guide the pixel sampling strategy in the rendering pipeline.

The scenes were engineered to contain a variety of conditional heuristics to adhere to: smell congruency and naturally more dominant salient features. The scenes are termed Lounge, Kitchen and Restaurant. Presenting participants with visual cues under different smell conditions, the experiment observes whether or not traditional saliency models of visual feature prediction can be added to for another sensory stimulus (olfactory) in a selective-rendering context.

5.2.1.2. Render Cost Function: In a selective-rendering pipeline given a map as a heuristic to weight the sampling strategy; a cost to compute an image in terms of the degree of sampling used can be given as:

$$V = \sum_{x=0}^{w}\sum_{y=0}^{h} s_{\min} + ((s_{\max} - s_{\min}) \cdot sal(x,y)) \tag{3}$$

where V is the number of samples required to compute an image, $s_{\min}$ is the minimum number of samples used to calculate radiance through a pixel, $s_{\max}$ is the most number of samples used to calculate this, $sal(x,y)$ is the weighting coefficient for a specific pixel in image space, and x and y are pixel coordinates.

Scene	Sampling	s_{min}	s_{max}	Avg. SPP
Lounge	Uni	1600	1600	1600
Lounge	Sal	1200	5000	1624
Lounge	SmSal	500	5000	1619
Kitchen	Uni	600	600	600
Kitchen	Sal	200	2105	617
Kitchen	SmSal	200	2000	607
Restaurant	Uni	75	75	75
Restaurant	Sal	70	90	74
Restaurant	SmSal	70	95	76

Table 5: Render Cost Function Across Scenes. Avg. SPP (Samples per pixel) is the average number of samples used to generate an image, dictating complexity.

To investigate the perceptual difference between two selective-rendering strategies it is necessary to control this render cost function so that, given a number of samples to generate each image, V, an optimisation process starts to vary s_{min} and s_{max} such that $(S \approx V) \pm f$ where f is some user defined control of sufficient leeway to compensate for the fact that s_{min} and s_{max} are restricted to integers in the optimisation process and S is the actual number of samples used in the generation of the image.

Varying render cost functions were used to investigate if the technique was applicable generically or not. Table 5 shows the various s_{min}, s_{max} for the various sampling methods used.

5.2.1.3. Olfactory Stimuli The smells were delivered through the Dale Air Vortex Activ emission system via micro controller. However smells are not standardised or controlled and one smell purporting to smell like one thing may not necessarily be the case. Presented olfactory stimuli were analysed using Gas Chromatography Mass Spectrometry (GCMS), the various synthetic Volatile Organic Compounds (VOCs) used in this experiment were compared to a collection of real VOCs to provide similarity data. The method used follows after the ISO standard developed and outlined in [War04]. This was done in the case of the lavender smell; the other smells delivered, bread and banana, also had their makeup analysed but are not presented here for brevity. The setup for collection of VOCs is shown in Figure 6.

The procedure used the micro-controlled emission system to deliver 6 minutes of volatiles through the collection piping. The flow rate at the end of the delivery systems PTFE tubes was recorded

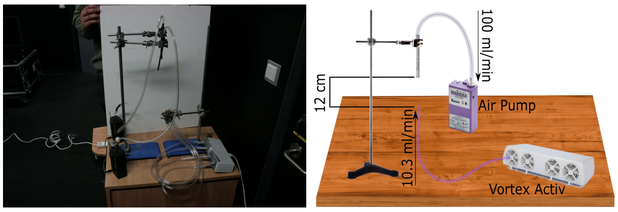

Figure 6: Real (l) and schematic (r) diagrams highlighting the collection mechanism for VOC's, synthetic and real, through the tenax tubing.

at 10.3 ml/min. This volatile compromised air was drawn through the collection tubes at a controlled rate of 100 ml/min, roughly 12 cm (the average distance in the experiments from the end of the PTFE tubes and a participant's nostrils) above the delivery tubes. For the real collection instead of the end of the delivery device 12 cm below the collection tubes a cutting of lavender was located into the clamp and the same procedure was followed of 6 minute exposure to the unsealed tenax tube. Once the collection was completed the tenax tube was sealed air tight with a clamping lock. The tenax tubes were parsed through GCMS. There was a similarity between real and synthetic measurements of all collected VOCs. For example, in the case of the lavender real and synthetic captures, there existed similar molecular abundance peaks of Toluene-d8 (9.785s), Butanamide (17.434s) and Phthalic acid, butyl isohexyl ester (27.815s).

5.2.2. Equipment

The equipment used in this experiment is the same as that presented in Section 2.2.3 minus the use of the eye tracker. The smell delivery system used was a Vortex Activ produced by Dale Air. This device has four fan emission chambers which are programmable via a micro controller to guide chamber impulse onset, duration and which chamber in the device to fire. PTFE tubes were used to minimise the effect of adsorption of smell molecules onto the surface of the inner tubing which could bias subsequent data collection as participants went through the experiment. Tubes were blown clear before and after each participant, as well as an empty chamber pumping air when a smell chamber was not firing. In addition to this, during experiment down time, room fans and a room neutral deodorant were used in order to minimise discrepancies between subsequent data sets. The resolution of the LCD panel was 1024×768 at 60 Hz and images were computed at this resolution.

5.2.3. Setup

Experimental setup and placement is shown in Figure 1. The participant was sat on a chair, with the backrest of the chair 115cm from the display. One metre long PTFE tubes were drawn from the device chambers and clipped to the top of a participants collar. Setup ensured that each participant's eyes were aligned with the centre of the screen, which was achieved by adjusting the height of the chair they were sitting on. The experiment was conducted in a dark room to avoid any effects of ambient lighting and participants were allowed time in order to adjust to the environment for 5 minutes before commencing the actual experiment. Slide presentation order and location (A or B) were randomised to remove any ordering bias. The time allowed to make a choice between the rendered images was 5 seconds which was decided after a pilot study to determine the time required to make a quality difference judgement accurately. It is important to allow the same amount of time for each participant and it should be long enough for the participant to make an informed choice without being able to over analyse. The participants were briefed prior to the experiment in order to gain a clear understanding of their task and the experiment did not proceed until this was certain.

5.3. Procedure

Image slides were presented synchronously with the smell cues for a total of five seconds slide exposure. Buffer slides, providing information as to the current slide were present for two seconds in between the image slides and the decision slide was present for a minimum of ten seconds to provide a buffer for nose desensitisation after olfactory stimulus. The experiment lasted an average of 12 minutes per participant. Smells used in this experiment were lavender, crusty bread and banana; the fourth chamber on the emission device was kept clear to provide a means to push through clean air into the nostrils. Smells were delivered to the participant for the full duration of the relevant slides. Figure 7 shows the ordering of the buffer and image slides used in the experiment and indicates slide timings. Table 6 highlights the different smells used on a per scene basis.

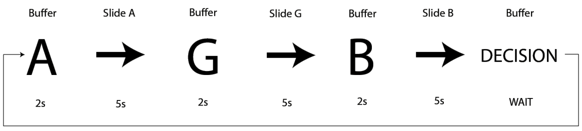

Figure 7: Experimental Image Slide Ordering.

Scene	Smell
Lounge	Lavender
Kitchen	Crusty Bread
Restaurant	Banana

Table 6: Smells used per scene

5.4. Participants

A total of 28 participants took part in this experiment, 21 males and 7 females. Selection criteria, participation and recruitment is the same as of that in Section 2.4. The age range of participants was between 21 and 42, with an average age of 27. Each participant was presented with all of the scenes, thus looking at a total of $\frac{t(t-1)}{2} \cdot 3 = \frac{3(2)}{2} \cdot 3 = 9$ image pairs.

5.5. Analysis

The analysis of the participants' judgements was performed with a *balanced design paired-comparison test* [KBS40], where each subject was instructed to evaluate all possible comparison pairs taken from the test set. This enables the evaluation and comparison of the performance of each test subject. t is given as the number of scenes rendered with different selective criteria that are to be compared against each other against the reference rendered image. For any given scene, each participant was presented $\binom{t}{2} = 3$ pairs $\frac{t(t-1)}{2}$, all possible combinations of rendering strategies to be compared.

For each comparison pair, the participant's choice was recorded. Once all of the pairs have been presented, the results from the participant's choices are recorded in a $t \cdot t$ matrix. An example of this matrix is shown in Table 7. Denote p_i as the number of preferences scored by Test i where $i \in 1, 2 \ldots t$, then the overall score per condition per subject is given as:

$$\sum_{i=1}^{t} p_i = \frac{t(t-1)}{2} = 3 \tag{4}$$

The votes for all s participants conducting the pairwise-comparisons are combined into a summed preference matrix per rendering condition. If all of the subjects completely agreed in their paired-comparisons, then $\frac{t(t-1)}{2}$ matrix entities would have a value of s and the remaining entities would have a value of 0. Note that the central diagonal is never considered since the same image is not pair compared against itself, yielding a value of 0 for the applicable section of the preference table.

	Test 1	Test 2	Test 3	Score
Test 1	*	0	0	0
Test 2	1	*	0	1
Test 3	1	1	*	2

Table 7: Example preference matrix for one subject when shown three rendered images of a given scene. Each rendered image is presented in a row, Tests in a row are compared with another Test in each column.

The overall similarity results for the 3 scenes are shown in Tables 8, 9 and 10. The paired-comparison data is provided in Table 12 and coloured rings highlight the perceptual grouping that exists in between the selective-rendering strategies on a per scene basis.

5.5.1. Statistical Analysis

p_{ij} is the number of times that an image i is preferred to image j by a participant. The sum of this result per participant, excluding the condition where $i = j$, is given as Σ:

$$\Sigma = \sum_{i \neq j}^{t(t-1)} \binom{p_{ij}}{2} \tag{5}$$

where, t is the number of selective-rendering strategies to be considered. Σ is the sum of the number of agreements between pairs.

For brevity, the names of the rendering strategies have been abbreviated in this manuscript. These names are Uniform (Uni), Visual Saliency (Sal) and Olfactory-Visual Saliency (SmSal).

Performing paired-comparisons requires analysis of participant *consistency*. When evaluating the three conditions (triad), selective-rendering strategies, Uni, Sal and SmSal in this experiment: the participant gives a judgement, for example, Uni > Sal, Sal > SmSal and Uni > SmSal. The judgement given by the participant is deemed to be *consistent* when the results given hold across the board, Uni > Sal > SmSal. Inconsistent choices such as SmSal > Uni introduces what is called a circular triad. Although inconsistency is not ideal in analytical scenarios, where a typical ranking approach is adopted it is necessary to handle participant inconsistencies due to tiredness or levels of perceived difference being small and a choice being difficult to make in the given time. In addition, inconsistency does not definitely mean that the data is wrong. It can provide statistical evidence that the images presented for p_{ij} are highly similar and the judgements being made are, as a result, inconsistent; whereas blindly ranking the choices wouldn't consider this conclusion. However, if the inconsistencies are present in only a small proportion of participants, then it can be concluded that they are not capable of making a consistent judgement whereas most other participants are, and then there exists a justification to not consider the determinations given by that particular participant. The coefficient of consistency is produced on a scale of $0 \cdots 1$ with 0 indicating complete inconsistency and 1 indicating complete consistence. A coefficient of consistency ζ is given by the equation:

$$\zeta = 1 - \frac{24c}{t^3 - 4t} \tag{6}$$

where c is the number of circular triads observed per participant per test condition. If $\zeta = 1$, there are no circular triads. The coefficient of consistency in this experiment was $\zeta_{average} \approx 0.71$ and as such the participant's consistency was deemed to be good and can all be included in the paired-comparison study. The number of circular triads is determined as follows [Dav69]:

$$c = \begin{cases} \frac{t}{24}(t^2 - 1) - \frac{1}{2}T, & t \mod 2 \equiv 1 \\ \frac{1}{24}(t^3 - 4t) - \frac{1}{2}T, & t \mod 2 \equiv 0 \end{cases} \tag{7}$$

$$T = \Sigma(p_i - \frac{t-1}{2})^2 \tag{8}$$

The Σ for each condition presented which have differences of less than $\pm R$ is deemed to not be significantly different and the conditions can be perceptual grouped into the same categories. However, the conditions with different perceptual groups are declared to be significantly different when $\Sigma \pm R$ does not fall in range with other values of Σ and the condition is awarded a separate perceptual grouping. The distribution of the range R is asymptotically the same as the distribution of the variance-normalised range, W_t, of a set of normal random variables with variance = 1 and t samples [Dav69]. Therefore, the following relation holds:

$$P\left(W_t \geq \frac{2R - \frac{1}{2}}{\sqrt{st}}\right) \tag{9}$$

where $W_{t,a}$ is the value of the upper percentage point of W_t at significance point $a = 0.05$. The values of $W_{t,a}$ can be obtained from statistics books, for example [PH88]. Given $W_{t,a}$, it is possible to find R':

$$R' = \frac{1}{2} W_{t,a} \sqrt{st} + \frac{1}{4} \quad (10)$$

The null hypothesis, H_0', is that all conditions are equal under testing ($H_0' : \pi_i = \frac{1}{2}$). The alternative being that not all the conditions π_i are equal. If the score difference for a given scene between two rendering conditions is larger than R^+ (the smallest integer greater than R'), the conclusion is that there is a statistically significant difference between the two conditions presented and this indicates that one is perceptually closer to the ideal reference image than the other. The multiple comparison range test has the property of making it difficult for true differences to show themselves. Yet the method allows comparisons to be performed after the initial inspection of experimental results and preference matrix generation. In addition, the probability of any incorrect declaration of grouping differences is controlled at the significance level reported in p. The preference tables for each scene are presented in Tables 8, 9 and 10.

	Uni	Sal	SmSal	Score
Uni	*	5	7	12
Sal	23	*	12	35
SmSal	21	16	*	37

Table 8: Preference matrix for the lounge scene presented with olfactory stimulus (lavender plant) and olfactory congruency. Rendering strategies used for image generation used as row and column headings.

	Uni	Sal	SmSal	Score
Uni	*	11	11	22
Sal	17	*	12	29
SmSal	17	16	*	33

Table 9: Preference matrix for the kitchen scene presented with olfactory stimulus (bread) and olfactory congruency. Rendering strategies used for image generation used as row and column headings.

	Uni	Sal	SmSal	Score
Uni	*	14	5	19
Sal	14	*	10	24
SmSal	23	18	*	41

Table 10: Preference matrix for the restaurant scene presented with olfactory stimulus (banana bunch) and olfactory congruency. Rendering strategies used for image generation used as row and column headings.

	Uni	Sal	SmSal	Score
Uni	*	30	23	53
Sal	54	*	34	88
SmSal	61	50	*	111

Table 11: Preference matrix for all of the scenes combined presented with corresponding olfactory stimulus and olfactory congruency. Rendering strategies used for image generation used as row and column headings.

6. Discussion

Table 8 shows close relationship between the Sal and SmSal operators and the Uni operator was weakest amongst this group in a direct comparison. SmSal was chosen in a paired-comparison over the Sal rendering technique more times, 16 to 12 whilst the Sal condition was evaluated higher than Uni than SmSal was with a score of 23 to 21. Table 9 shows equal comparisons to Uni in both cases of Sal and SmSal, however again SmSal is chosen more than Sal in that pairwise test with a score of 16 to 12. Table 10 shows a large discrepancy in results; Uni and Sal are both chosen much less frequently than the SmSal image and SmSal is again chosen over Sal, 18 to 10. Table 11 shows the studies for each scene combined into one preference matrix.

The results provided an R^+ (the smallest integer greater than R') of 18. This is shown against cumulative votes for each scene and aggregated in Figure 8. This resulted in the following groupings shown in Table 12. As can be seen the results were perceptually indistinguishable statistically in the Kitchen scene coupled to which the null hypothesis is not rejected as $\chi^2_{df=2,p<0.05} = 5.99 > 3.444$. In the case of the Restaurant scene and Lounge scene perceptually identifiable groups emerged. In both scenes where there was ambiguity, the Sal operator has been chosen fewer times over the SmSal condition in a direct pairwise choice, there were fewer circular triads. So in addition to coming at the top of the preference tables the SmSal operator had fewer discrepancies and was perceptually distinguishable statistically in one scene and in the overall test cases. In order to analyse the data in an aggregated fashion, the scenes have also been collapsed to enable evaluation across the scenes combined. This increased the overall power of the analysis, but ignores the dependency on scene. However, it is clear to see in Table 12 that significant groupings appeared for each of the rendering strategies in this *Overall* case. This shows, that in the general case, this strategy is beneficial over both Uniform and Saliency metrics for guiding sampling strategy in a VE alongside an olfactory stimulus.

7. Conclusions and Future Work

This paper has presented a method which seems to exploits the Human Visual System's (HVSs) *bottom-up* approach. The fact that the HVS is guided by other modal impulses allows selectively rendering the regions attended to in higher quality and the remainder of the image in a lower quality capitalising on inattentional blindness. The results extend previous work of smell and graphics, such as [BCB*09a] and confirms the impact the inclusion of smell into

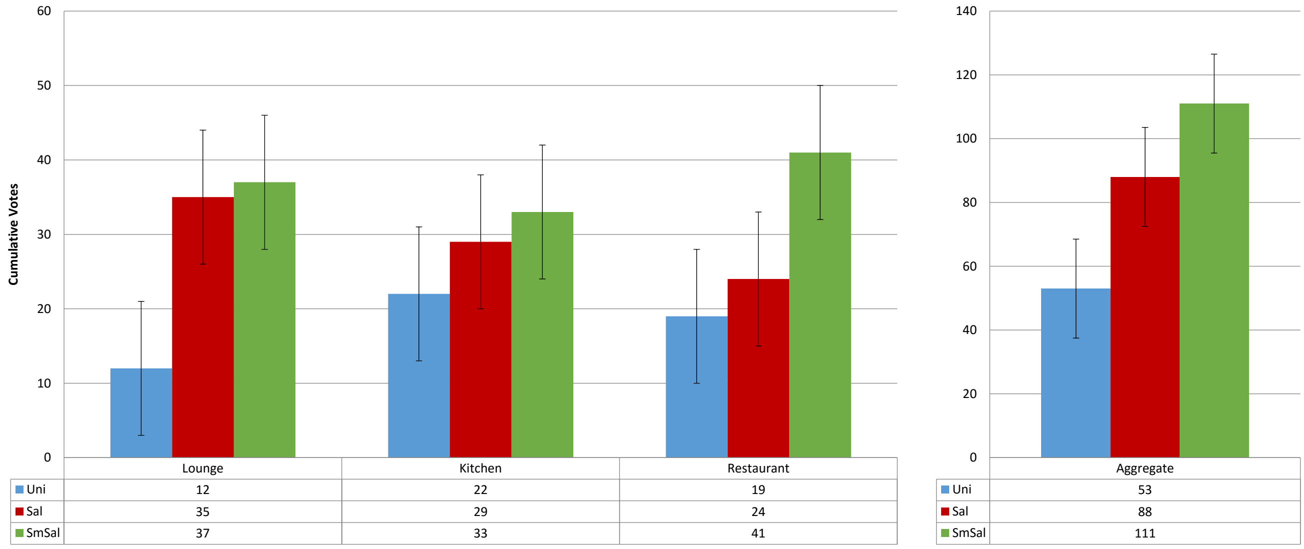

Figure 8: Method Preference; Error bars indicate the range R^+. R^+ is 18 for individual scenes and 31 for the aggregate.

	ζ	Kendall's W	χ^2	P, df=2	Rank 1	Rank 2	Rank 3
Lounge	0.7143	0.300	16.783	<0.05	SmSal*	Sal	Uni
Kitchen	0.6786	0.062	3.444	=0.179	SmSal*	Sal	Uni
Restaurant	0.7500	0.226	12.667	<0.05	SmSal	Sal	Uni
Overall	0.7143	0.336	18.808	<0.05	SmSal	Sal	Uni

Table 12: Overall similarity study conclusion for the various scenes presented with olfactory stimuli and olfactory congruency's. *SmSal is ordered first in the ranking from placements in the preference table but it is worth noting the results contain fewer circular triads than the Sal condition.

high-fidelity VEs has for selective-rendering. Smell is a key human sense. Including olfaction into VEs improves perceived realism.

Future work will investigate the variability of the weighting function w across variable smell sets. In evolutionary terms, certain smells are more prevalent to certain people, and in fact; some demographs are anosmic or less responsive to some smells. The hedonic tone of a smell is highly subjective as it is related to both genetic and environmental factors. This makes it difficult to categorise smells as positive or negative smells. The effects described here have not been tested for participants with a degree of anosmia to specific smells. In addition it is necessary to investigate the effect multiple smell impulses have on visual attention. A similar avenue of research is to study the intensity of the smell impulse at which the effect on visual attention comes into play; whilst the human nose can detect odours subliminally in parts per billion, the threshold to which this effect comes into play must lie between the exposure the experiments presented in this chapter used and a baseline reading. It is important to determine where this threshold lies and whether it is variable and quantifiable between various smell types from an odour labelling perspective. Finally, a regression algorithm will be investigated to find the value of w that fits the olfactory-visual saliency map to the observed gaze tracking data retrieved during the experiment. This investigation will provide further insight into the inner workings of olfactory-visual perceptual interactions.

References

[Act10] ACTIV V.: Dale air ltd. World Wide Web electronic publication, http://www.daleair.com/, 2010. 3

[AMMG16] ADEMOYE O. A., MURRAY N., MUNTEAN G.-M., GHINEA G.: Audio masking effect on inter-component skews in olfaction-enhanced multimedia presentations. *ACM Trans. Multimedia Comput. Commun. Appl. 12*, 4 (Aug. 2016), 51:1–51:14. 1

[BCB*09a] BRKIC B., CHALMERS A., BOULANGER K., PATTTANAIK S., COVINGTON J.: Cross-modal affects of smell on real-time rendering of grass. In *In SCCG'09: Spring Conference on Computer Graphics* (2009), ACM SIGGRAPH Press, pp. 175–179. 10

[BCB*09b] BRKIC B. R., CHALMERS A., BOULANGER K., PATTANAIK S., COVINGTON J.: Cross-modal affects of smell on the real-time rendering of grass. In *Proceedings of the 25th Spring Conference on Computer Graphics* (New York, NY, USA, 2009), SCCG '09, ACM, pp. 161–166. 1

[Bla97] BLAUERT J.: *Spatial Hearing : The Psychophysics of Human Sound Localization*. M.I.T. Press, Cambridge, MA, 1997. 1

[BSVDD10] BONNEEL N., SUIED C., VIAUD-DELMON I., DRETTAKIS G.: Bimodal perception of audio-visual material properties for virtual environments. *ACM Trans. Appl. Percept. 7*, 1 (2010), 1–16. 1

[CCL02] CATER K., CHALMERS A., LEDDA P.: Selective quality rendering by exploiting human inattentional blindness: looking but not seeing. In *VRST '02* (New York, NY, USA, 2002), ACM, pp. 17–24. 1

[CZC*13] CHEN K., ZHOU B., CHEN S., HE S., ZHOU W.: Olfaction spontaneously highlights visual saliency map. *Proceedings of the Royal Society of London B: Biological Sciences 280*, 1768 (2013). 2

[Dal] DALENBÄCK B.-I.: CATT-Acoustic, Gothenburg, Sweden. www.netg.se/catt. 1

[Dav69] DAVID H. A.: *The Method of Paired Comparisons*. Charles Griffin and Company, London, 1969. 9

[Deb06] DEBATTISTA K.: *Selective Rendering for High Fidelity Graphics*. PhD in Computer science, University of Bristol, 2006. 6

[DM95] DURLACH N., MAVOR A.: *Virtual Reality Scientific and Technological Challenges*. Tech. rep., National Research Council Report, National Academy Press, 1995. 1

[GBW*09] GRELAUD D., BONNEEL N., WIMMER M., ASSELOT M., DRETTAKIS G.: Efficient and practical audio-visual rendering for games using crossmodal perception. In *I3D '09* (New York, NY, USA, 2009), ACM, pp. 177–182. 1

[GTLG14] GHINEA G., TIMMERER C., LIN W., GULLIVER S. R.: Mulsemedia: State of the art, perspectives, and challenges. *ACM Trans. Multimedia Comput. Commun. Appl. 11*, 1s (Oct. 2014), 17:1–17:23. 1

[HBRDC11] HARVEY C., BASHFORD-ROGERS T. E. W., DEBATTISTA K., CHALMERS A.: Visual Saliency for Smell Impulses and Application to Selective Rendering. In *TPCG'11, Theory and Practice of Computer Graphics* (2011), Grimstead I., Carr H., (Eds.), Eurographics Association, pp. 73–80. 1, 2

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis, 1998. 2

[iLa] ilab neuromorphic vision c++ toolkit:. Available: http://ilab.usc.edu/toolkit/. 3

[Jac07] JACOB T.: A tutorial on the sense of smell. World Wide Web electronic publication, http://www.cardiff.ac.uk/biosi/staffinfo/jacob/teaching/sensory/ olfact1.html, 2007. 1

[Kaj86] KAJIYA J. T.: The rendering equation. *SIGGRAPH Comput. Graph. 20*, 4 (Aug. 1986), 143–150. 6

[KBS40] KENDALL M., BABINGTON-SMITH B.: On the method of paired comparisons. *Biometrika 31* (1940), 324–345. 9

[KC14] KATSUKI F., CONSTANTINIDIS C.: Bottom-up and top-down attention. *The Neuroscientist 20*, 5 (2014), 509–521. 2

[Ken75] KENDALL M.: *Rank Correlation Methods*, 4 ed. Griffin, 1975. 6

[Kna95] KNASKO S. C.: Pleasant odors and congruency: Effects on approach behavior. *Chem Senses 20*, 5 (1995), 479 – 487. 2

[MBM02] MILLOT J. L., BRAND G., MORAND N.: Effects of ambient odors on reaction time in humans. *Neurosci Lett. 322*, 2 (Apr 2002), 79 – 82. 2

[MBT*07] MOECK T., BONNEEL N., TSINGOS N., DRETTAKIS G., VIAUD-DELMON I., ALLOZA D.: Progressive perceptual audio rendering of complex scenes. In *I3D '07: Proceedings of the 2007 symposium on Interactive 3D graphics and games* (New York, NY, USA, 2007), ACM, pp. 189–196. 1

[MDCT05] MASTOROPOULOU G., DEBATTISTA K., CHALMERS A., TROSCIANKO T.: Auditory bias of visual attention for perceptually-guided selective rendering of animations. In *GRAPHITE '05* (New York, NY, USA, 2005), ACM Press, pp. 363–369. 1

[Nay93] NAYLOR J.: ODEON - another Hybrid Room Acoustical Model. *Applied Acoustics 38*, 1 (1993), 131–143. 1

[PH88] PEARSON E., HARTLEY H.: *Biometrika Tables for Statisticians*, 3 ed. Cambridge University Press, 1988. 10

[RBF08] RAMANARAYANAN G., BALA K., FERWERDA J. A.: Perception of complex aggregates. In *SIGGRAPH '08: ACM SIGGRAPH 2008 papers* (New York, NY, USA, 2008), ACM, pp. 1–10. 1

[RCHR07] RAMIC B., CHALMERS A., HASIC J., RIZVIC S.: Selective rendering in a multi-modal environment: Scent and graphics. In *Proceedings of the 23rd Spring Conference on Computer Graphics* (New York, NY, USA, 2007), SCCG '07, ACM, pp. 147–151. 1

[RFWB07] RAMANARAYANAN G., FERWERDA J., WALTER B., BALA K.: Visual equivalence: towards a new standard for image fidelity. *ACM Trans. Graph. 26*, 3 (2007), 76. 1

[RKH*13] ROJAS D., KAPRALOS B., HOGUE A., COLLINS K., NACKE L., CRISTANCHO S., CONATI C., DUBROWSKI A.: The effect of sound on visual fidelity perception in stereoscopic 3-d. *IEEE Transactions on Cybernetics 43*, 6 (Dec 2013), 1572–1583. 1

[RLC*07] RAGHUVANSHI N., LAUTERBACH C., CHANDAK A., MANOCHA D., LIN M. C.: Real-time sound synthesis and propagation for games. *Commun. ACM 50*, 7 (2007), 66–73. 1

[SC88] SIEGEL S., CASTELLAN N.: *Nonparametric Statistics for the Behavioral Sciences*. McGrall-Hill International, 1988. 6

[SDJ*10] SEIGNEURIC A., DURAND K., JIANG T., BAUDOUIN J., SCHAAL B.: The nose tells it to the eyes: crossmodal associations between olfaction and vision. *Perception 39*, 10 (2010), 1541 – 1554. 1, 2

[SIS*05] SAKAI N., IMADA S., SAITO S., KOBAYAKAWA T., DEGUCHI Y.: The effect of visual images on perception of odors. *Chem Senses 30*, 1 (2005), 244 – 245. 1

[SRMN10] SEO H.-S., ROIDL E., MÜLLER F., NEGOIAS S.: Odors enhance visual attention to congruent objects. *Appetite 54*, 3 (Jun 2010), 544 – 549. Epub. 2, 5

[TGD04] TSINGOS N., GALLO E., DRETTAKIS G.: Perceptual audio rendering of complex virtual environments. *ACM Trans. Graph. 23*, 3 (2004), 249–258. 1

[War04] WARWICK UNIVERSITY ISO STANDARD, BS ISO16200-1: Workplace air quality: sampling and analysis of volatile organic compounds by solvent desorption/gas chromatography, 2004. 7

[YPG01] YEE H., PATTANAIK S., GREENBERG D. P.: Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics 20*, 1 (January 2001), 39–65. 7

[ZBE91] ZELLNER D. A., BARTOLI A. M., ECKARD R.: Influence of color on odor identification and liking ratings. *The American Journal of Psychology 104*, 4 (1991), 547 – 561. 1, 2